

2

AD-A256 419

AD \_\_\_\_\_



ANTIVIRAL DRUGS: MOLECULAR MODELING AND QSAR

FINAL REPORT

KEITH B. WARD

DECEMBER 10, 1990

DTIC  
ELECTE  
OCT 20 1992  
S C D

Supported by

U.S. ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND  
Fort Detrick, Frederick, Maryland 21702-5012

Army Project Order No. 88PP8815

Naval Research Laboratory  
4555 Overlook Avenue, S.W.  
Washington, DC 20375-5000

Approved for public release; distribution unlimited.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents

82 1 19 039

251900

92-27326



88  
195

## REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

1a REPORT SECURITY CLASSIFICATION Unclassified			1b RESTRICTIVE MARKINGS		
2a SECURITY CLASSIFICATION AUTHORITY			3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited		
2b DECLASSIFICATION/DOWNGRADING SCHEDULE					
4 PERFORMING ORGANIZATION REPORT NUMBER(S)			5. MONITORING ORGANIZATION REPORT NUMBER(S)		
6a. NAME OF PERFORMING ORGANIZATION Naval Research Laboratory		6b. OFFICE SYMBOL (If applicable)		7a. NAME OF MONITORING ORGANIZATION	
6c. ADDRESS (City, State, and ZIP Code) 4555 Overlook Avenue, S.W. Washington, DC 20375-5000			7b. ADDRESS (City, State, and ZIP Code)		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Medical Research & Development Command		8b. OFFICE SYMBOL (If applicable)		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER Project Order No. 88PP8815	
8c. ADDRESS (City, State, and ZIP Code) Fort Detrick Frederick, Maryland 21702-5012			10. SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO. 63002A	PROJECT NO. 3M2- 63002D807	TASK NO. AD
			WORK UNIT ACCESSION NO. DA314675		
11 TITLE (Include Security Classification) Antiviral Drugs: Molecular Modeling and QSAR					
12 PERSONAL AUTHOR(S) Keith B. Ward					
13a. TYPE OF REPORT Final		13b. TIME COVERED FROM 3/1/88 TO 4/14/89		14. DATE OF REPORT (Year, Month, Day) 1990 December 10	
15. PAGE COUNT					
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP			
20	02		Drug discovery; QSAR; Antiviral drugs; Atom		
06	13		pairs method; RA 1		
19. ABSTRACT (Continue on reverse if necessary and identify by block number)					
<p>In order to provide guidance for a USAMRIID drug development program, an assessment was made of the major needs of USAMRIID in this field. Methods and techniques were implemented to fulfill these requirements. Because of the paucity and lack of quality of objective assesment of antiviral activity of compounds being explored, a major effort involved developing novel methodology that could deal successfully with these data. Only when LSM workers were granted free access to computerized data bases developed for USAMRIID by Techna Associates was progress forthcoming on the effort.</p> <p>Software was developed using the Atoms Pairs method to describe the topology of compounds under investigation. This has produced a method to identify lead compounds, to identify cogenes to active compounds not in the Technq database and to identify new types of compounds not previously investigated for antiviral activity.</p>					
20 DISTRIBUTION/AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/IN LIMITED <input checked="" type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS			21 ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a. NAME OF RESPONSIBLE INDIVIDUAL Mary Frances Bostian			22b TELEPHONE (Include Area Code) 301-619-7325		22c. OFFICE SYMBOL SGRD-RMI-S

## FINAL REPORT

### Antiviral Drug Discovery

Project Order No. 88PP8815  
Covering 01 March 88-14 April 89

## INTRODUCTION

The goal of our efforts was to assist the antiviral drug development program at U.S.A.M.R.I.I.D. (USAMRIID) for the development and design of antiviral drugs. During this contract these efforts have included:

- (1) evaluation of the needs of USAMRIID and how we could best fulfill our goals;
- (2) implementation of the methods and techniques that we deemed of greatest use to fulfill those needs; address, wherever possible, specific issues that could further USAMRIID's progress; aid USAMRIID in specific projects and in answering specific questions related to drug design and our areas of expertise;
- (3) provide general education to USAMRIID personnel in our methods and capabilities.

In terms of drug design effort, the antiviral drug development effort at USAMRIID is in its infancy. Little is known of the structure or biology of the viruses in question and little is known of the mechanisms of action of the antiviral drugs. A drug development project is composed of many stages and can be approached from several directions. One direction is from an understanding of the mechanism of the disease state to be treated.

Unfortunately because of the dearth of information about the viruses, for the antiviral work this approach to drug design is currently unavailable to USAMRIID and to us. USAMRIID is interested in as many as 18 viruses and for none of these is there detailed information regarding structure, composition, or biological machinery. Another approach to drug design is more empirical and consists of an evaluation of the chemical structure and biological activity of current drugs as well as other compounds that have been evaluated for the desired activity. This is the only approach that is available to us until more is known about these viruses.

In an effort to evaluate the antiviral activity and relate it to chemical structure, we have explored, extended, and utilized a method of evaluating the topological similarity of compounds. This method allows for the unbiased and automatic screening of large numbers of chemical compounds in order to evaluate them for their similarity or dissimilarity to compounds that have already been studied, very likely compounds that possess promising antiviral activity.

During a previous contract, (see **Appendix A**, containing **Final Report** for that effort, which contains useful background information for the appreciation of this current report) we utilized the antiviral activity of known or candidate antiviral drugs in order to develop a crude index of antiviral activity of known drugs based on their topological similarity to those drugs. However, during recent consultations with the researchers at USAMRIID, they indicated that they were much less confident of many of the biological activities provided by the in vitro screening than they had been during our previous contract. Because of this, during the past year our applications of the

topological similarity screening concentrated on methods and approaches that were dependent not on strict interpretations of biological activities but rather on qualitative estimates of activity or no estimates at all. This required reevaluation of our approaches and consultations with USAMRIID to determine the best course of action, which we describe below.

This last year is the first that we have had unhampered access to the computerized chemical database developed by Techna, Associates. This database contains chemical structures and the antiviral activity of those structures against several viruses, in several assays, supplied by several laboratories. In order to access the data stored in that database, Terry Stouch received instruction in the use of the MACCS chemical database computer software that Techna used to store and organized this information. He was also granted an account on Techna's DEC MicroVAX computer.

### **Evaluation of the Problem**

A drug's activity depends on many factors, including route of administration, route of transport through an organism to the site of action, retention time in the body, metabolism and detoxification, and the extent of the particular reaction or interaction at the actual site of action of physiological effect. A drug design effort must take all of these factors into account. As we mentioned above, the best approach to drug design is through detailed knowledge of drug's fate in all of these steps and possibly others. This knowledge is very difficult and expensive to acquire, however, and over the last several decades many other ways have been proposed to guide the drug design effort. These methods include, random screening Quantitative Structure Activity Relationship studies (QSAR), analysis of molecular topology, pattern recognition, molecular modeling, and applications of quantum chemistry.

We have considered each of these methods for application to USAMRIID's problems. USAMRIID's current approach is effectively one of directed random screening of compounds. This method consists of screening a random selection of compounds in various assays for the desired antiviral activity. This screening is said to be directed because prior to assay the compounds are evaluated by knowledgeable personnel. USAMRIID's current method of selecting the compounds that are screened is not efficient and is not really random, however. As described below, some of our efforts are directed towards increasing the efficiency of USAMRIID's random screening efforts.

As we describe below, QSAR approaches require series of related compounds. Work during our previous contract showed that USAMRIID does not yet have such series for their most promising antivirals. Molecular modeling and quantum chemical approaches are time consuming and are usually performed on one or only a small number of compounds at any one time. Since USAMRIID's effort involves thousands of compounds, these approaches are currently impractical. Pattern recognition methods require some homology of structure or mechanism in order to be most effective. Since USAMRIID is interested in identifying antiviral agents to many different viruses and, the current known antiviral drugs are structurally diverse, pattern recognition can not yet be effectively applied to USAMRIID'S's current most pressing needs.

After evaluation of the antiviral drug development project, the time frame of our efforts, and the facilities at both USAMRIID and the NRL, we decided that the best approach to aiding USAMRIID, at least initially, is to apply methods of comparing the topological structure of molecules. Our decision was based on the fact that these methods

1. are computationally of reasonably cost,
2. could be implemented in a reasonable amount of time
3. yield results quickly,
4. can be used only large numbers of compounds (this is a very important feature of the studies that we performed and recommend be performed, as will be elaborated on in this report)
5. do not require 3-dimensional structures and
6. have been proven effective in drug design and discovery efforts.

These methods can be used to help identify interesting new compounds and series of related compounds that can be studied using the the other, more detailed methods of analysis mentioned above. The topological methods could assist USAMRIID in their random screening, in exploiting known antivirals, and in determining new lead compounds

### Methods

By molecular topology we mean the atoms and their elemental types that are within a molecule and the ways that the atoms are connected. Such information is easy to handle with computer and is contained within the structure of a molecule, as it is normally perceived. Many methods of describing molecular topology have been proposed (eg. see References by Hodes, Carhart, Johnson). Many are somewhat arbitrary in their description of topology and omit sometimes useful information and some describe molecular topology using an essentially endless list of structural features. One method that combines many of the best features of the others and does not suffer from some of the drawbacks is the Atom Pairs method as described by Carhart et al. (Carhart, 1985). This method generates a limited number of features, is unambiguous, uses all of the information within the chemical structures being described and, most importantly, has shown considerable success in other drug design efforts.

Briefly the Atom Pairs method works by dividing a molecular structure into its constituent pairs of atoms. Information regarding distance between atoms, elemental type of the atoms, and valence of the atoms is used to develop an "atom pair." A particular molecule is represented in the computer by a list of all of the atom pairs that it contains. Two molecules can be compared and contrasted by directly comparing their lists.

If the list are equivalent, then the two molecules contain the same atom pairs with the same number of occurrences of each and the molecule are identical. If they have no atom pairs in common, then they are completely different. Many numerical indexes can be used to described this comparison. We have chosen to use a method suggested by Carhart et al. (Carhart, 1985) that yields a continuous series of numbers from 0.0 to 1.0 where a value of 0 means that the molecules have nothing in common and a value of 1.0 means that they are identical. A more complete description of both the atom pairs method and the index described above can be found in Carhart, 1985 and Nilakantan, 1987, and Jaeger, 1984.

This method can make thousands of comparisons within minutes and can look through

large computerized databases for compounds that are similar to or different from a target compound. Carhart also suggested using biological activity data to develop 'trend vectors' mathematical constructs that would relate molecule based both on their constituent atom pairs and on their biological activities.

As mentioned previously, during a previous contract, we developed our own method of combining biological activity information with the atom pairs formalism. We developed software to calculate biological activity weights for each atom pair contained in a list of all atom pairs that were contained within all the molecules of interest. We then used those weights and a molecule's list of constituent atom pairs to calculate a crude index of biological activity. This method was successful at differentiating active from inactive compounds when applied to hundreds of compounds of known antiviral activity. It can also be used to calculate a crude index of activity for new, untested compounds.

### Approaches

As mentioned in the introduction, the apparent poor quality of the antiviral assay data led us to investigate approaches to aiding USAMRIID that were not dependent on evaluation or correlation of that data. Given that constraint, we thought that we could best do this by applying the topological search method to:

1. aid in the identification of new lead compounds based on the structures of known antivirals
2. identify congeners to active compounds that are not in the U/Techna database and so derive the most benefit from that data and
3. identify new types of compounds that have not been previously investigated for antiviral activity.

Each of these points will be addressed in turn, followed by our efforts to act on them.

Points 1 and 2: The key axiom of the field of Structure-Activity Relationship (SAR) studies is that a molecule's chemical structure is responsible for its physical and chemical properties and that those properties are responsible for its biological activity. Variation in chemical structure will cause variation in the biological activity. The field of SAR aims to derive the relationships between the variations. A corollary to this axiom is that compounds that are similar to each other will have similar biological activities. This corollary guides our efforts into the implementation of points 1 and 2, above. We sought to find compounds that were similar to known antivirals at two levels, identified by the two separate points, 1 and 2. Point 1 targets the identification of new lead compounds. A lead compound is one that possess the desired biological activity but appears structurally dissimilar from other classes of compounds with that same activity. For example, ribavirin, a promising antiviral compound and a nucleoside analog, would represent a lead compound. Compounds from other distinct chemical classes, such as those of steroids or hydrazoles, would represent other lead compounds. Minor changes to lead compounds by small variations in structure (single atom addition or replacement) in order to increase activity or reduce unwanted side effects, represents lead optimization. The techniques embodying this latter step comprise, in the strictest sense,

Quantitative Structure Activity Relationship Studies (QSAR) about which much has been written and discussed.

In general then, Point 1 addresses the task of finding compounds similar to those with known antiviral activity but which do not have strict homology of structure. Point 2 addresses the task of finding congeners to those known compounds, compounds that differ only slightly from those with promising activity profiles.

The similarity of chemical compounds is often difficult to determine. Congeners are usually easy to spot. However, sometimes the chemical features responsible for the biological activity of a compound is deeply imbedded within a structure and is difficult to determine. That is why the Atom Pairs method of determining similarity is useful. It is an exhaustive, unbiased, automatic method of comparing structures. It reduces compounds to their smallest contiguous fragments and compares two structures at that level. In this way, related structures can be discovered that would have been missed by even highly trained human researchers. This is particularly true since a computer based method can search many thousands of compounds exhaustively whereas a human has difficulty even with several tens of compounds.

Point 1: In addressing point 1, we came to the conclusion that the atom pairs software should be used to screen large computerized databases of compounds in a search for compounds that represented new leads. There are several reasons for this. First, the current method being used by USAMRIID to obtain new leads involves soliciting compounds from academia, industry, and government labs and then having a panel of scientist visually screen them. This method suffers in several respects. First, they will not get a representative sampling of compounds; many will be similar to each other or will represent intermediates along a synthetic pathway. Second, they will have only a finite number of compounds to examine. Third, the capability of human screening of compounds for activity is very dependent on many factors including experience, and physical and time demands.

Many computerized chemical databases are available that contain thousands of compounds. By interfacing the ATOM PAIRS method of similarity searching, all of these tens of thousands of compounds could be search systematically and exhaustively in an unbiased manner.

Point 2: Point 2 is also addressed in much the same way. While congeners to known antivirals are much easier to spot, it is impossible for a human, in a reasonable period of time, to examine hundreds of thousands of compounds. Even the use of computer-aided substructural screening for congeners is difficult to do exhaustively. The implementation of similarity screens, such as we propose, will speed and ease this process. By rapidly identifying congeners, these compounds can be assayed for activity. The resulting series of compounds can then be analyzed in a detailed way to fully exploit the information within the active compounds to learn in a detailed way, at the atomic level, about the structural features affecting activity.

Point 3: Point 3 consists of performing the opposite procedure to that used in addressing points 1 and 2, however, it uses the same techniques. USAMRIID is interested in determining the identity of ANY potential antiviral drug, regardless of structural type. Points 1 and 2 addressed determining the structures of compounds that were similar in some way to those compounds that we already know have high activity. However, as we stated above, the range and breadth of the structural types and functionality of those compounds is limited. In addition to exploring similar compounds, USAMRIID is also interested in investigating completely novel compounds. For this reason, we have investigated ways and methods to use the similarity index to identify compounds that have novel features.

Such features could be a completely different structural type, a chemical functional group not present in the compounds already in the data base, or an elemental type not present in the database. Such new features will most likely result in a different chemistry that might not be present in the already existing database of compounds.

All three points above could be addressed by interfacing our atom pairs software system with some large computerized repository of chemical structure information. Many such repositories exist in industry, government, and some are available commercially. Such repositories contain computer-readable chemical structures of known compounds, many contain information on chemical and physical properties of these compounds, some contain biological information such as toxicities, and others contain information on synthetic routes or suppliers. The compounds contained in these databases could be searched for compounds that conform to the requirements of points 1,2, and 3, above.

We concluded that for point 1, the structural information contained in the compounds with the most promising antiviral profile should be encoded and used in the search through the database in an effort to find new lead compounds. For point 2, the database could be searched for congeners to those most promising compounds. Careful evaluations of the antiviral activity of these congeners could be followed by detailed QSAR studies in an effort to optimize the activity of that series and understand their mechanism of antiviral activity. For point 3, only very different, completely novel compounds would be identified. These would be brought to USAMRIID attention as candidates for experimental antiviral activity screening in an effort to bring some system to the discovery of new compounds.

After the tasks of evaluating the problems and determining suitable approaches for their solutions, much of our effort has been spent in determining:

1. the best way to conduct the searches suggested above,
2. the appropriate databases to examine and
3. the necessary software modifications needed in order to access those databases, and
4. other hardware and software requirements of these searches.

### Searches

The results of our examination of the similarity indexes of 463 compounds provided to us from their database by Techna, Associates. indicate that best way to conduct these searches is to apply three different approaches to address the three separate searches.

For determination of new lead compounds (Point 1) that, although being new lead compounds, still contain structural similarities to known antivirals, we conclude that selecting new compounds which exhibit a similarity index of between 0.4 and 0.6 will supply compounds that show some structural resemblance to the known compounds while not being so similar as to



constitute congeners. Indexes of higher value will identify compounds with close structural similarities, those with lower values identify compounds with little or no similarity. In order to allow the maximum amount of flexibility in this search and in order to reduce computational requirements, we have determined a two tiered evaluation approach could be used. First, a composite list of the atom pairs contained within all of the known antivirals should be developed and used for comparison with new compounds. Those new compounds showing few or no occurrences of the atom pairs contained in that list should not be considered. The second tier consists of a detailed molecule-by-molecule comparison of each of the new compounds that passed the first with each of the known antivirals. Within that second tier, any compounds with similarity indexes within the range noted above to any or all of the known antivirals should be further examined as possible new lead compounds.

For determination of congeners of known antivirals (Point 2), a detailed compound-by-compounds comparison of the known antivirals to all of the compounds in the searched database should be undertaken. Any compound exhibiting a similarity index of 0.95 or above to a known antiviral would constitute a congener. These compounds should be acquired and assayed for antiviral activity since they represent a minor structural variation on a compound of promise. Obtaining such biological information and incorporation of these compounds into the Techna database will provide needed information for QSAR studies and hence for the most efficient exploitation of the structural/biological information inherent in the known antiviral.

For determination of compounds from completely new classes (Point 3) a composite list of atom pairs should be compiled for ALL of the compounds that have been already examined. Any compound containing none of these atom pairs, any containing a new element, or any compound which is composed of 50% or less of known atom pairs should be experimentally screened for biological activity because it would represent an entirely new type of compound that had not yet been examined. A more detailed compound-by-compound comparison of each of the compounds in the searched database to each of the compounds in the Techna database would be very time consuming. If such a comparison was undertaken, compounds with a similarity index of less than 0.1 to ANY of the compounds in the Techna database should be examined.

### Selection of Databases

Several computerized databases of chemical structures are available and include MedLine, Toxline, the National Cancer Institute maintains a large database, many have been developed in private industry, particularly in the pharmaceutical industry. The Institute for Scientific Information provides a database commercially. Each of these has advantages and disadvantages. Some are very large and contain hundreds of thousands of compounds. Some contain information on biological activities such as toxicity. Others are more readily available than others.

We investigated the possibility of using several of these and decided that, at least initially, the Fine Chemicals Directory was most promising. There are several practical reasons for our conclusion. First, this database contains compounds that are commercially available. It would be convenient to obtain compounds that the similarity screening identified as of interest. Compounds in other databases might not be so readily available and synthesis might be required. While USAMRIID has contracts for synthetic work in place, such synthesis could require months or years, especially if tens or hundreds of compounds are involved. A rapid response would be best both for USAMRIID interests and our own. We still consider these similarity screens to be experimental, the information gained through its application will be valuable in judging its effectiveness and in helping to refine its use. Second, while not the largest of the databases, the

FCD does contain many tens of thousands of compounds. Third, the FCD has been put into MACCS data files. MACCS is the database management system used by Techna, Associates to organize U data. This will make the FCD data more convenient to access than that in most of the other databases. We are only suggesting that this effort start with the FCD. As more experience is gained with the application of these methods, large databases, perhaps those containing biological activity information for the compounds they contain, can be searched.

### Mechanics of Implementation

In order to perform the proposed searches, the atom pairs software must be provided with the connectivity data for the compounds of interest. Such connectivity data consists of the atoms present in the molecule, their elemental type, which other atoms each atom is connected to, and the bond type of those connections. To date, our investigations into the use of these similarity screens have been performed in the environment of the ADAPT chemical software system. (ADAPT, Automated Data Analysis and Pattern recognition Toolkit, is a suite of computer programs designed specifically for analysis of chemical and SAR problems that Dr. Terry Stouch helped to design when he was a graduate student at The Pennsylvania State University. ADAPT is marketed by Molecular Design, Ltd., San Leandro, CA.).

ADAPT was the most convenient system for us to use because, due to his work on this software system and his affiliations, Dr. Stouch has access to all the source code and data file formats for that software. Much of the atom pairs similarity index generating software was written using the libraries of ADAPT FORTRAN subroutines for chemical data handling. There were many practical aspects that we needed to address in creating an interface, a software link, between the atom pairs / ADAPT software and the MACCS database where the FCD structures were stored. Previously, Techna had used MACCS to output structures in a form that was readable by the ADAPT system. This was somewhat cumbersome, but adequate for the several hundreds of compounds that we had been dealing with.

This is an impractical method for dealing with the 70,000 plus compounds in the FCD, however. The ADAPT interface would no longer be practical either, since in its current form, the ADAPT data files can hold only 1000 compounds at one time. We investigated methods of expanding ADAPT, but found that the data size limitations were inherent in the overall software architecture. Such modifications would require extensive revision and much time. Another alternative that we explored was to interface directly with the MACCS database and extract the needed structural information for the FCD compounds directly. We discovered that the Molecular Design Limited, the vendor of MACCS, considers the MACCS database data storage format to be proprietary and so this avenue was unavailable to us. Our conclusion was that the most practical and expedient procedure would be to extract all the compounds from the FCD database (in MACCS format) and output them in the intermediate format that was used previously to transfer the compounds to the ADAPT system. The atom pairs software would be converted to "stand alone" without support from the ADAPT software system. These conversions would allow it to directly read the intermediate data files.

This conversion could be substantial, since in reading the intermediate files, the ADAPT system applies some intelligence to interpreting some of the chemical bonding information (such as aromaticity, dative bonds, ionic bonds). This work requires three to six months of the efforts of a dedicated scientific programmer skilled in FORTRAN and DEC VAX systems and programming. In addition to the conversions, mentioned

above, a programmer could be of value in refining the experimental atom pairs software into a system that would be more automatic and convenient for use by researchers at USAMRIID and Techna, Associates.

## CONSULTATIONS

As during our previous contract, in this last year we continued to advise the USAMRIID antiviral drug development program on topics concerning quantitative structure activity relationships and related computational tools as applied to drug design.

During several visits to USAMRIID, we discussed the possibility of examining the antiviral activity data within the Techna database for trends and correlations between the different viral assays. If any of the assays is strongly correlated with one or several other assays, then that assay would provide only redundant information. Since this information is expensive and time-consuming to acquire, such a redundant assay could be assigned a low priority or discontinued. Trends within the data could provide information on the similarity of mechanism of antiviral action of drugs between the different viruses. Indirectly, this could provide information on similarities in the biochemistry of the different viruses.

Unfortunately, as mentioned above, the available biological activity data was not consistent, and in some cases it was not sound. Some of the data was of higher quality and more reliable than others. At the conclusion of our consultations concerning these studies, it was decided that the USAMRIID researchers would provide us with a set of "rules" which we could use to interpret and evaluate the biological activity data as well the data that they were most interested in evaluating. With this data, we were to conduct the pertinent data analysis. Unfortunately, the data and the information that we required were not made available to us.

We had several consultations with Bjørn Gabrielson concerning possible quantitative structure-activity relationship studies. One of these concerned series of S-adenosyl-methionine (SAM) hydrolase inhibitors (SHI). SHI is implicated in the biochemistry of many viruses. Modification of the function of this enzyme could be a powerful tool for affecting virus livelihood. Another consultation involved evaluating methods of analyzing the structure-activity relationships of a series of 20 steroid-like, compounds containing a lactam fused to an aromatic ring (Figure 1). Dr. Gabrielson was interested in investigating the application of three-dimensional approaches of studying the structural features as they related to antiviral activity. The most extensive consultation with Dr. Gabrielson involved a study of the SAR of a series of 25 substituted benzoxy adenosine compounds (Figure 2), many of which had high activity against vaccinia virus. These compounds and their reported antiviral activities were obtained from Techna and entered into ADAPT data files (ADAPT, Automated Data Analysis and Pattern recognition Toolkit, is a suite of computer programs designed specifically for analysis of chemical and SAR problems that Dr. Terry Stouch helped to design when he was a graduate student at The Pennsylvania State University. ADAPT is marketed by Molecular Design, Ltd., San Leandro, CA.). Preliminary analysis of this data was begun.

Prior to the attainment of any conclusions, however, this project was placed at low priority by the COTR of this grant.

### CONCLUSIONS:

In conclusion, we have determined approaches that we recommend be used by USAMRIID for enhancing their antiviral drug design project. We determined methods that will aid their effort by both discovering new lead compounds as well as more fully exploiting the information contained within those compounds already exhibiting promising antiviral activity. We recommend that they search other larger chemical databases against the data in their own database to attain these goals and we identified the database that not only is the most convenient one to use at the present time from a practical standpoint but also should have the breadth and scope that this antiviral drug development project needs and, in addition, should provide compounds that are convenient to acquire. We also determined the practical and computer programming steps that need to be taken in order to perform the searches we recommend on the database that we recommend.

Finally, whenever requested we have rendered aid to USAMRIID personnel in assisting them with SAR problems and in evaluating the promise of interesting series of compounds.

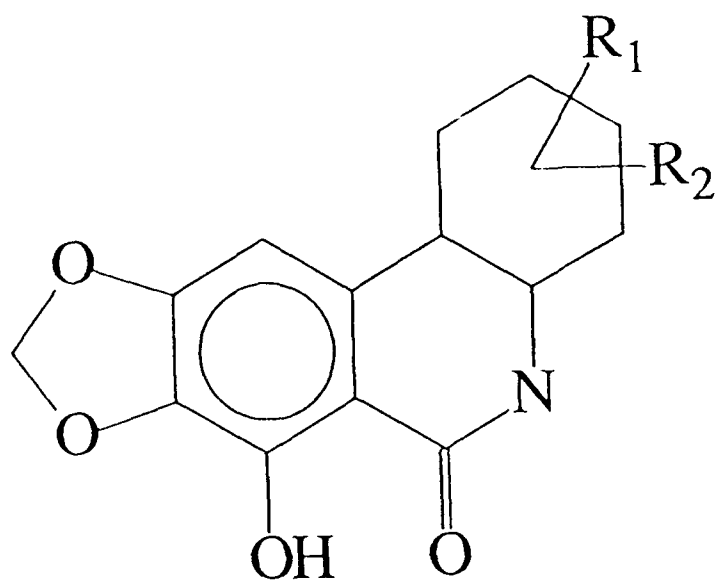
### References

- Carhart, R., Smith, D.I., Venkataraghavan, R., *J. Chem. Inf. Comput. Sci.*,  
1985, **25**, 64-73.
- Hodes, L., *J. Med. Chem.*, **20**(4), 1977, 469-475.
- Hodes, L., *J. Chem. Inf. Comput. Sci.*, **21**, 128-132.
- Hodes, L., *J. Med. Chem.*, **29**, 2207-2212.
- Jaeger, E., Jurs, P., in "Analytical Chemistry Instrumentation," W.R. Lang,  
Ed., Lewis Publishers, Chelsea, Michigan, 1986.
- Johnson, M., Naim, M., Nicholson, V., Tsai, C. in "QSAR in Drug Design and  
Toxicology," Elsevier, 1987.
- Johnson, M. and Basak, S., Proceedings of the 6th International Conference  
on Mathematical Modeling, Saint Louis, August 4-7 1987.

Nilakantan, R., Bauman, N., Dixon, J., Venkataraghavan, R., *J. Chem. Inf. Comput. Sci.*, 27(2), 1987, 82-85.

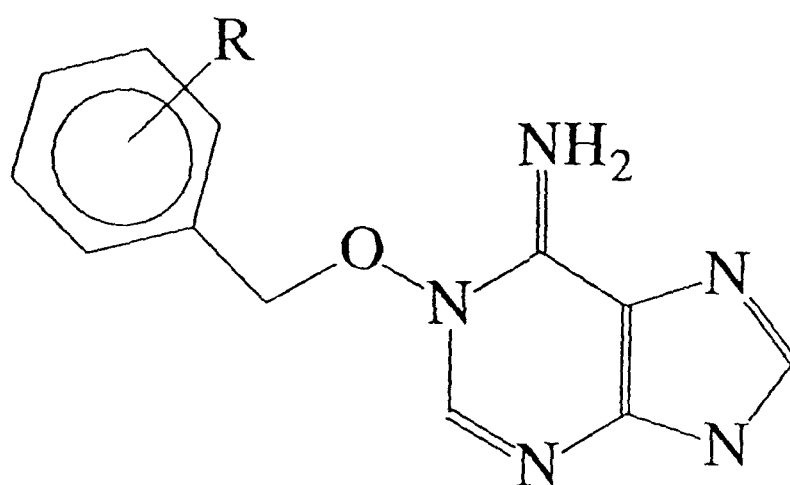
Willett, P., Winterman, V., Bawden, D., *J. Chem. Inf. Comput. Sci.* 27(2) 1987, 109-117.

**Figure 1**



**Figure 1:** Structural backbone of a series of steroidal-lactams.

**Figure 2**



**Figure 2:** Structural backbone of the Benzoyloxy adenosine compounds.

## Appendix A

### Final Report on Antiviral Drug Discovery Component of Previous Contract

#### INTRODUCTION:

The goal of our efforts was to assist the antiviral drug development program at U.S.A.M.R.I.I.D. (USAMRIID) for the development and design of antiviral drugs. During this contract these efforts have included evaluation of the needs of USAMRIID and implementation of the methods and techniques that we deemed of greatest use to fulfill those needs. These methods and techniques included some directed at aiding the prioritization of the testing of compounds and some directed at helping to most fully exploit information contained in active antiviral drugs.

In terms of drug design effort, the antiviral drug development effort at USAMRIID is in its infancy. Little is known of the structure or biology of the viruses in question and little is known of the mechanisms of action of the antiviral drugs. A drug development project is composed of many stages and can be approached from several directions. One direction is from an understanding of the mechanism of the disease state to be treated.

Unfortunately because of the dearth of information about the viruses, for the antiviral work this approach to drug design was unavailable to USAMRIID and to us. USAMRIID was interested in as many as 18 viruses and for none of these is there detailed information regarding structure, composition, or biological machinery. Another approach to drug design is more empirical and consists of an evaluation of the chemical structure and biological activity of current drugs as well as other compounds that have been evaluated for the desired activity. This is the only approach that was available to us at the beginning of our contract and will continue to be the only approach until more is known about these viruses.

A drug's activity depends on many factors, including route of administration, route of transport through an organism to the site of action, retention time in the body, metabolism and detoxification, and the extent of the particular reaction or interaction at the actual site of action of physiological effect. A drug design effort must take all of these factors into account. As we mentioned above, the best approach to drug design is through detailed knowledge of drug's fate in all of these steps and possibly others. This knowledge is very difficult and expensive to acquire, however, and over the last several decades many other ways have been proposed to guide the drug design effort. These methods include, random screening Quantitative Structure Activity Relationship studies (QSAR), analysis of molecular topology, pattern recognition, molecular modeling, and applications of quantum chemistry.

We considered each of these methods for application to USAMRIID's problems. USAMRIID's approach was effectively one of directed random screening of compounds. This method consists of screening a random selection of compounds in various assays for the desired antiviral activity. This screening is said to be directed because prior to assay the compounds are evaluated by knowledgeable personnel. USAMRIID's method of selecting the compounds that are screened is not efficient and is not really random, however. As described below, some of our efforts were directed towards increasing the efficiency of USAMRIID's random screening efforts.



As we describe below, QSAR approaches require series of related compounds. Our work showed that USAMRIID does not yet have such series for their most promising antivirals. Molecular modeling and quantum chemical approaches are time consuming and are usually performed on one or only a small number of compounds at any one time. Since USAMRIID's effort involved thousands of compounds, these approaches were impractical for our use. Pattern recognition methods require some homology of structure or mechanism in order to be most effective. Since USAMRIID was interested in identifying antiviral agents to many different viruses and the current known antiviral drugs are structurally diverse, pattern recognition could not be effectively applied to USAMRIID's most pressing needs.

After evaluation of the antiviral drug development project, the time frame of our efforts, and the facilities at both USAMRIID and the NRL, we decided that the best approach to aiding USAMRIID, at least initially, was to apply methods of comparing the topological structure of molecules. Our decision was based on the fact that these methods

- 1) Have shown demonstrated utility at Lederle Laboratories (Carhart, et al. 1985, Carhart, R. E., personal communications).
- 2) Could be implemented at the Naval Research Laboratory in a reasonable amount of time.
- 3) Have, as their major utility "lead" generation, which is the phase in which U.S.A.M.R.I.I.D. was, and continues to be, involved.
- 4) It is easily applied to large data bases, if those data bases are properly formatted. If such formatting is not present, further software development must precede its use.
- 5) Could be applied to preexisting data. No compound synthesis or testing is required.
- 6) Does not rely on potentially ambiguous estimates of antiviral activity.

## Methods

By molecular topology we mean the atoms and their elemental types that are within a molecule and the ways that the atoms are connected. Such information is easy to handle with computer and is contained within the structure of a molecule, as it is normally perceived. Many methods of describing molecular topology have been proposed (eg. see References by Hodes, Carhart, Johnson). Many are somewhat arbitrary in their description of topology and omit sometimes useful information and some describe molecular topology using an essentially endless list of structural features. One method that combines many of the best features of the others and does not suffer from some of the drawbacks is the Atom Pairs method as described by Carhart et al. (Carhart, 1985). This method generates a limited number of features, is unambiguous, uses all of the information within the chemical structures being described and, most importantly, has shown considerable success in other drug design efforts.

Briefly the Atom Pairs method works by dividing a molecular structure into its constituent pairs of atoms. Information regarding distance between atoms, elemental type of the atoms, and valence of the atoms is used to develop an "atom pair." A particular molecule is represented in the computer by a list of all of the atom pairs that it contains. Two molecules can be compared and contrasted by directly comparing their lists. If the list are equivalent, then the two molecules contain the same atom pairs with the same number of occurrences of each and the molecule are identical. If they have no atom pairs in common, then they are completely different. Many numerical indexes can be used to described this comparison. We have chosen to use a method suggested by Carhart et al. (Carhart, 1985) that yields a continuous series of numbers from 0.0 to 1.0 where a value of 0 means that the molecules have nothing in common and a value of 1.0 means that they are identical. A more complete description of both the atom pairs method and the index described above can be found in Carhart, 1985 and Nilakantan, 1987, and Jaeger, 1984.

An illustration of the individual structural descriptors, the atom pairs, that were calculated for one compound, ribavirin, is shown in Appendix 1. These atom pairs consist of information about each atom in the pair and the distance (as number of bonds in the shortest path ) between these atoms. The information encoded for each atom includes: 1) the elemental type, 2) the number of "pi" electrons, and, 3) the number of nonhydrogen attachments. For example, the atom pair with the index of '1' is the first atom pair in the list. The two atoms in the pair are separated by 10 bonds, hence the "10" between the atom descriptions which are in the parenthesis. The atom description of the leftmost atom shows that it is an oxygen (the first character within parenthesis) with no "pi" electrons, and one attachment to a nonhydrogen atom. The atom description on the right indicates a nitrogen atom with no "pi" electrons, and one attachment. This atom pair shows that there is within this molecule a hydroxyl group that is 10 bonds away from a primary amine function. The "Frequency" column shows the number of times that this atom pair occurs in ribavirin (once). The "Key" column is a numerical encoding of the atom pair that is convenient for internal computer representation.

An illustration of the information that the similarity about the compounds of a data set is shown in Figure 1. Here, ribavirin is shown along with 32 other compounds. The leftmost number near the top of each structural graph is the structure number used by the ADAPT structure files and has no chemical significance. ADAPT is an integrated software system for performing

structure-activity studies and the atom pairs software has been interfaced to it. The handwritten number to the right of the structure number is the similarity index for each compound relative to ribavirin (ADAPT structure number 1). This index is scaled so that it will assume values between 0 and 1. A compound that is very similar to another will have a high index value relative to the latter compound. A very dissimilar compound will have a value approaching 0.

Within Figure 1, ribavirin is identical to itself and so has a similarity index of 1. The first compound to the right of ribavirin, compound 48, has a similarity to ribavirin of 0.941. It can be seen from the chemical graph that this compound is structurally very similar to ribavirin; it differs only in the exchange of two atoms (circled). All the compounds with similarity indexes greater than 0.88 are structurally very similar to ribavirin and differ by only one atom or the exchange of two atoms. Compounds with an index of 0.80 to 0.84 differ by two atoms. As the structures become less similar, the similarity index drops, also. Compound 413, the last compound listed in Figure 1, has an index value of 0.012 and can be seen to be quite different from ribavirin.

### Software

A large effort on our part involved acquisition, conversion, implementation, debugging, and testing the computer software required for these studies.

### Biological Activity Indicator

In addition to the similarity index, the atom pairs descriptors have been used to generate a crude indicator of biological activity for new, untested compounds. This indicator is similar to the statistical-heuristic method of Hodes. In both methods, the topological graphs of a large number of compounds of known activity are reduced to subgraphs, in our case atom pairs. These subgraphs are then assigned activity weights based on the activities of the compounds that they occur within. For example, if a subgraph is found only within compounds of high activity, then its active weight will be high and its non-active weight will be small. If a subgraph occurs equally within the active and inactive classes, its active and non-active weights will be equal.

The activity of an untested compound can be estimated by the appropriate summation of the weights of the subgraphs that the chemical graph of that compound contains. There are a variety of methods for calculating and evaluating the weights, and we are currently involved in the estimation of several of these.

The results of one method are illustrated in Figure 2. The data used to generate this plot was that of the Rift Valley Fever (RVF) therapeutic index (TI) for 462 compounds supplied by Techna Associates. This particular test was chosen because our collaborators at Fort Detrick are most interested in this test and also because of the paucity of data for the other seven viruses tested. Even for the RVF test, only 14 compounds had TIs of greater than 50, a value that Techna Associates considers the cutoff between active and inactive. Of the 462 compounds, over 70 had no assay performed at all.

Figure 2 was generated as follows. The atom pairs and atom pairs weights were calculated for the 390 compounds that had a reported TI value for the RVF assay. An estimate of the activity of each of the 462 compounds supplied by Techna was then calculated based on these

weights. This was done by determining the atom pairs within the structure, summing the active and non-active weights for those atom pairs and dividing the summation of the active weight by the non-active weight (which was negative). The actual activity of each compound was then plotted versus this index. A low value for this ratio indicates a high estimated activity. This plot shows good discrimination for the active compounds. Over the range of values, from -0.2 to -0.55, the bulk of the compounds with TIs of greater than 50 had estimated activity values of -0.44 or less, indicating high activity. The two exceptions were compounds of lower activity and were still significantly above the low end of the range.

The inactive compounds are distributed over the entire range of values. This is thought to be an artifact of the means of calculating the weights. That region of the plot most densely populated by the inactive compounds was at a higher ratio value than that area most densely populated by the active compounds, however.

Based on plots like this, we can make suggestions as to which of a potentially large number of compounds to be assayed first if, in fact, all of the compounds can not be assayed. For example, the compounds with the value of -99 did not have any reported activity. If these were to be assayed, this plot would indicate that those with a ratio of less than -0.4 would have the greatest chance of being biologically active.

We show this plot as an example and note that there is little information contained within 14 active compounds as compared to 376 inactive compounds. In order to confidently use the results of such plots, more information must be supplied.

Those activity estimation studies made use of the therapeutic indexes (TIs) of the antiviral compounds that are contained in the antiviral database that was available to us. TIs are determined by dividing an estimate of the toxicity of a compound by an estimate of its antiviral activity (ID50). Later work centered on the use of the ID50s of the compounds for activity estimation. This was done both in order to further assess the advantages of the activity estimation calculations and also to develop a method for estimating the ID50s. The ID50s might provide a more rigorous test of this methodology than do the TIs because 1) they are a less qualitative measure of activity than are the TIs and 2) the ID50s of the compounds are more uniformly distributed over a wider range than are the TIs. Furthermore, the ID50s are used routinely by USAMRIID in their drug screening program.

More detailed studies of the estimates of antiviral activity based on the activity weights were conducted in order to evaluate the predictive capabilities of this method and in order to determine the best approach of calculating these estimates. Table 1 shows results of the activity estimation for the Rift Valley Fever virus ID50 data. The activity of each of the compounds that were available to us and that were evaluated using this assay was predicted based on its constituent atom pairs. These estimates were then evaluated in reference to the actual activities. Table 1 shows the mean and standard deviation of the estimated activities of 1) all the compounds evaluated, 2) those compounds experimentally determined to have little or no activity and 3) those compounds experimentally determined to have appreciable antiviral activity. The means of the active and inactive classes of compounds were compared using the Student's T test. It can be seen that the difference between the means of the active and inactive compounds is highly significant at the 99.5% probability level. Several methods of calculating the activity estimates were investigated and appear as different lines in Table 1. All methods gave highly significant results with the method that we have encoded as "rcnt2" giving the greatest difference between active and inactive compounds.

What Table 1 shows is that this method yields a crude estimate of antiviral activity. Active compounds tend to have higher values of this index and inactive compounds tend to have lower values. Similar results were observed for estimates of activity for another virus, Japanese encephalitis. These results indicate that this methodology could be of use to the USAMRIID workers as a screen of new compounds for antiviral activity.

These results show this method to be internally consistent since estimates of the activities of the compounds that were used to develop the activity weights for the atom pairs (and which were used to calculate the estimated activities) showed highly significant differences between active and inactive compounds. Such consistency is a necessary but not sufficient requirement to guarantee that such methods will be useful in a truly predictive sense (to correctly predict the activity of compounds that were not used to calculate the activity weights). A series of studies was undertaken to assess such the true predictive capabilities of this method.

In the first, the activities that were used to calculate the RVF activity weights of the individual atom pairs were scrambled. This removed any real activity information that the resulting weights contained in the previous study. The activities of the compounds were again estimated as before. Statistical tests showed no significant difference between the estimated activities between the active and inactive classes. This verified that the difference noted in the last report was real, due to the biological information of the compounds, and not due to artifacts of the analysis or chance.

In the second study, inadvertently four of the most active compounds had not been included in the analysis of the RVF data. The estimated activities of these four were calculated according to the activity weights of their atom pairs. These estimates correctly indicated high activity. This is true evidence of the predictive capability of this method for antiviral activity, at least within the range of compounds and compound types included in this study.

### **Lead Optimization**

Our applications of this methods to a search of the database for congeners of active antivirals indicates that there are few analogues to some of the most promising antivirals: ribavirin and 4' esters of ribavirin. Both ribavirin and some of its 4' esters show high antiviral activity, however, there was no evidence in the data available to us that they have been properly exploited through sequential variations in their physical and chemical properties. Such variations can be used to try to determine compounds with enhanced activity or decreased negative side effects. Such variations can also be used to try to understand the mechanism of action and to compare and contrast the mechanism of action of different compounds. We recommend that these compounds be exploited more fully by a planned course of synthesis and testing of analogues.

### **Conclusions**

We have evaluated USAMRIID's needs and have suggested and tested methods that could be of value to them. These methods include topological similarity screening and crude predictors of antiviral activity. Our work has shown that these methods could be of value to their effort. We have evaluated that portion of the USAMRIID/Techna database that was made available to us and have made suggestions about the most fruitful avenues to proceed in a drug design sense. We

recommend that they implement these methods and use them routinely in their candidate drug evaluation process.

### References

- Carhart, R., Smith, D.I., Venkataraghavan, R., *J. Chem. Inf. Comput. Sci.* 1985, **25**, 64-73.
- Hodes, L., *J. Med. Chem.*, **20**(4), 1977, 469-475.
- Hodes, L., *J. Chem. Inf. Comput. Sci.*, **21**, 128-132.
- Hodes, L., *J. Med. Chem.*, **29**, 2207-2212.
- Jaeger, E., Jurs, P., in "Analytical Chemistry Instrumentation," W.R. Lang, Ed., Lewis Publishers, Chelsea, Michigan, 1986.
- Johnson, M., Naim, M., Nicholson, V., Tsai, C. in "QSAR in Drug Design and Toxicology," Elsevier, 1987.
- Johnson, M. and Basak, S., "Proceedings of the 6th International Conference on Mathematical Modeling", Saint Louis, August 4-7 1987.
- Nilakantan, R., Bauman, N., Dixon, J., Venkataraghavan, R., *J. Chem. Inf. Comput. Sci.*, **27**(2), 1987, 82-85.
- Willett, P., Winterman, V., Bawden, D., *J. Chem. Inf. Comput. Sci.*, **27**(2), 1987, 109-117.

### Figure Captions

**Figure 1:** Atom pairs similarity indexes of ribavirin (compound 1) to select other compounds in the USAMRIID database.

**Figure 2:** Actual Rift Valley Fever therapeutic index vs. estimated value of this index from the Atom Pairs activity weights.

Table 1

Activity Prediction from Atom Pairs Activity Weights

Rift Valley Fever

ID50 Data

	All			Inactive			Active			
method	mean	s.d.	N	mean	s.d.	N	mean	s.d.	N	t
<hr/>										
rcnt2	.370	.038	462	.361	.037	289	.393	.032	94	6.3
rfrq2	.380	.041		.371	.040		.401	.036		5.1
rfrq	.445	.061		.434	.061		.471	.060		7.7
rcnt	.417	.053		.405	.051		.443	.051		6.7
<hr/>										
f			for dof=381 at alpha=.005 t=2.6							

FIGURE 1

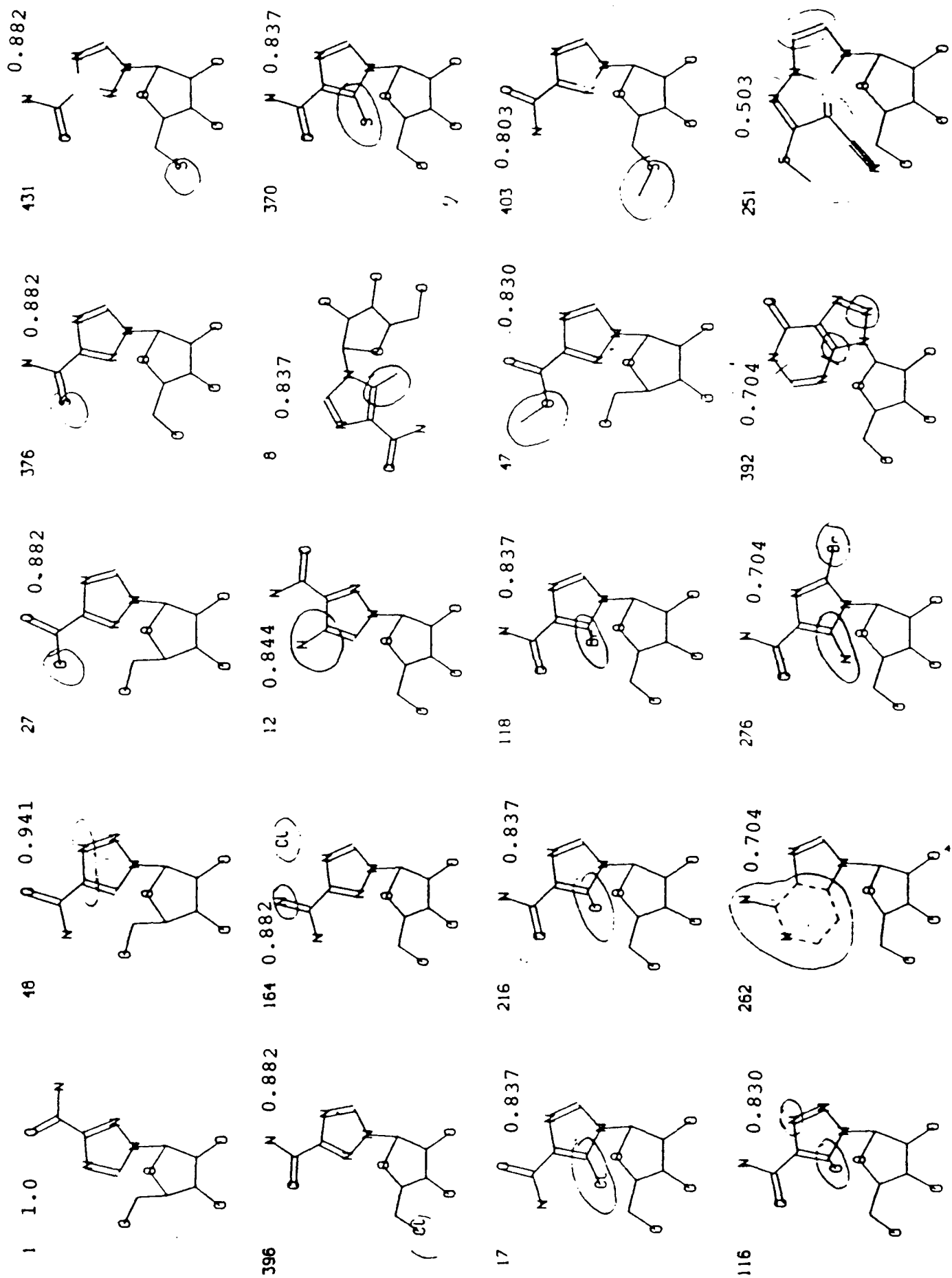




FIGURE 1 Continued

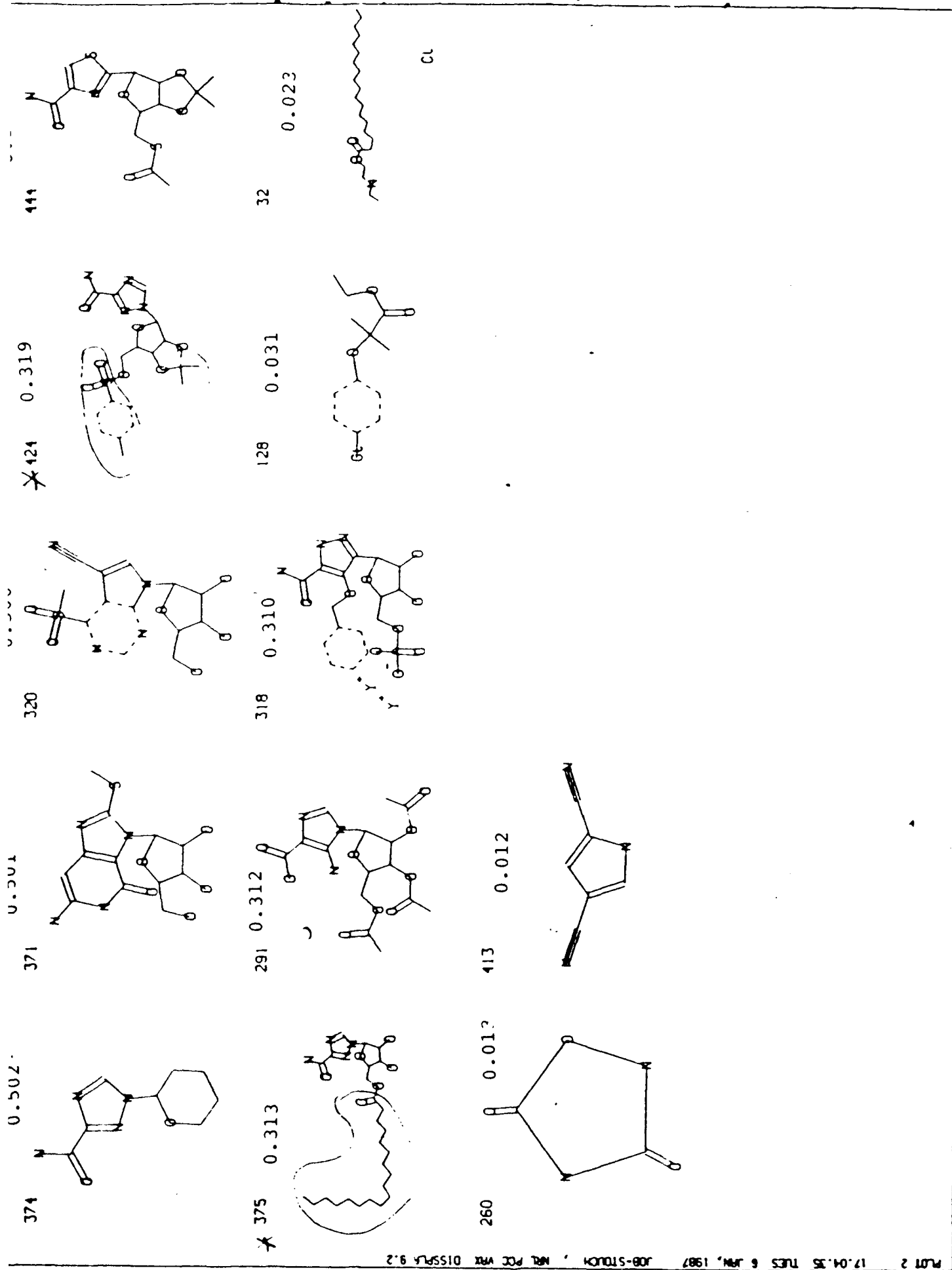
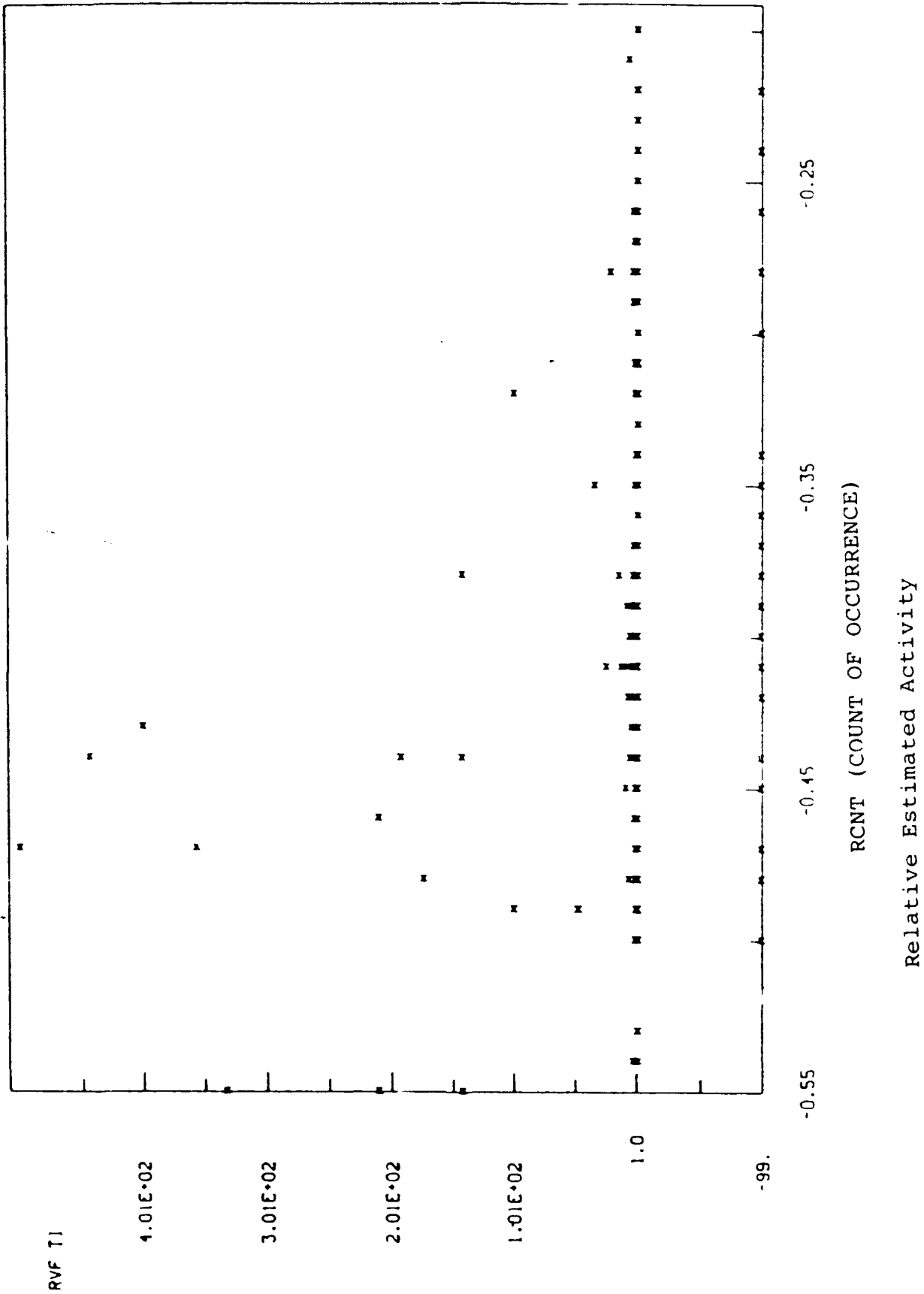


FIGURE 2

ATOM PAIRS WEIGHTED ACTIVITY ESTIMATION RVF TI VS R.F. RCNT



DAN= 1      NUMBER OF ATOM PAIRS= 136      NUMBER OF UNIQUE ATOM PAIRS= 99

INDEX	ATOM PAIR	FREQUENCY	KEY
1	(O :0:1) ---10--- (N :0:1)	1	10683521
2	(O :0:1) ---10--- (O :1:1)	1	10634369
3	(O :0:1) --- 9--- (N :0:1)	1	9634945
4	(C :0:2) --- 9--- (N :0:1)	1	9634882
5	(O :0:1) --- 9--- (O :1:1)	1	9585793
6	(C :0:2) --- 9--- (O :1:1)	1	9585730
7	(C :1:3) --- 9--- (O :0:1)	1	9569363
8	(O :0:1) --- 8--- (N :1:2)	1	8603777
9	(O :0:1) --- 8--- (N :0:1)	1	8586369
10	(C :0:3) --- 8--- (N :0:1)	2	8586307
11	(O :0:1) --- 8--- (O :1:1)	1	8537217
12	(C :0:3) --- 8--- (O :1:1)	2	8537155
13	(C :1:3) --- 8--- (O :0:1)	2	8520787
14	(C :0:2) --- 8--- (C :1:3)	1	8473666
15	(O :0:1) --- 7--- (N :1:2)	2	7555201
16	(C :0:2) --- 7--- (N :1:2)	1	7555138
17	(O :0:2) --- 7--- (N :0:1)	1	7537794
18	(C :0:3) --- 7--- (N :0:1)	1	7537731
19	(O :0:2) --- 7--- (O :1:1)	1	7488642
20	(C :0:3) --- 7--- (O :1:1)	1	7488579
21	(C :1:3) --- 7--- (O :0:1)	2	7472211
22	(C :1:2) --- 7--- (O :0:1)	1	7472210
23	(C :0:3) --- 7--- (C :1:3)	2	7425091
24	(C :0:2) --- 7--- (C :1:3)	1	7425090
25	(O :0:1) --- 6--- (N :1:2)	2	6506625
26	(C :0:3) --- 6--- (N :1:2)	2	6506563
27	(C :0:2) --- 6--- (N :1:2)	1	6506562
28	(O :0:1) --- 6--- (N :0:3)	1	6481265
29	(C :0:3) --- 6--- (N :0:1)	1	6489155
30	(C :0:3) --- 6--- (O :1:1)	1	6440003
31	(C :1:3) --- 6--- (O :0:2)	1	6424659
32	(O :0:1) --- 6--- (O :0:1)	1	6423681
33	(C :1:3) --- 6--- (O :0:1)	1	6423635
34	(C :1:2) --- 6--- (O :0:1)	1	6423634
35	(C :0:3) --- 6--- (C :1:3)	3	6376515
36	(C :0:2) --- 6--- (C :1:2)	1	6375490
37	(O :0:2) --- 5--- (N :1:2)	1	5458050
38	(O :0:1) --- 5--- (N :1:2)	1	5458049
39	(C :0:3) --- 5--- (N :1:2)	3	5457987
40	(N :0:1) --- 5--- (N :0:3)	1	5442753
41	(O :1:1) --- 5--- (N :0:3)	1	5442705
42	(O :0:1) --- 5--- (N :0:3)	1	5442689
43	(C :0:2) --- 5--- (N :0:3)	1	5442626
44	(C :1:2) --- 5--- (N :0:1)	1	5440594
45	(C :1:2) --- 5--- (O :1:1)	1	5391442
46	(C :1:3) --- 5--- (O :0:2)	1	5376083
47	(O :0:1) --- 5--- (O :0:1)	1	5375105
48	(C :1:2) --- 5--- (O :0:1)	1	5375058
49	(C :0:3) --- 5--- (O :0:1)	2	5375043
50	(C :0:2) --- 5--- (O :0:1)	1	5375042
51	(C :0:3) --- 5--- (C :1:3)	2	5327939
52	(C :0:3) --- 5--- (C :1:2)	2	5326915
53	(N :0:1) --- 4--- (N :1:2)	2	4409537
54	(O :1:1) --- 4--- (N :1:2)	2	4409489
55	(O :1:2) --- 4--- (N :1:2)	2	4409450

56	(C :0:3)	---	4---	(N :1:2)	2	4409411
57	(O :0:1)	---	4---	(N :0:3)	1	4394113
58	(C :1:3)	---	4---	(N :0:3)	1	4394067
59	(C :0:3)	---	4---	(N :0:3)	2	4394051
60	(O :0:1)	---	4---	(O :0:2)	3	4327553
61	(C :1:2)	---	4---	(O :0:2)	1	4327506
62	(O :0:1)	---	4---	(O :0:1)	1	4326529
63	(C :0:3)	---	4---	(O :0:1)	3	4326467
64	(C :0:2)	---	4---	(O :0:1)	1	4326466
65	(C :1:2)	---	4---	(C :1:3)	1	4279378
66	(C :0:3)	---	4---	(C :1:3)	1	4279363
67	(C :0:3)	---	4---	(C :1:2)	1	4278339
68	(C :0:2)	---	4---	(C :0:3)	2	4262978
69	(N :1:2)	---	3---	(N :1:2)	1	3360978
70	(N :0:3)	---	3---	(N :1:2)	1	3360963
71	(C :1:3)	---	3---	(N :1:2)	2	3360851
72	(C :1:2)	---	3---	(N :1:2)	1	3360850
73	(C :0:3)	---	3---	(N :1:2)	1	3360835
74	(O :0:2)	---	3---	(N :0:3)	1	3345538
75	(C :1:3)	---	3---	(N :0:3)	1	3345491
76	(C :0:3)	---	3---	(N :0:3)	1	3345475
77	(O :1:1)	---	3---	(N :0:1)	1	3343505
78	(C :1:3)	---	3---	(N :0:1)	1	3343443
79	(C :1:3)	---	3---	(O :1:1)	1	3294291
80	(C :0:3)	---	3---	(O :0:2)	2	3278915
81	(C :0:2)	---	3---	(O :0:2)	1	3278914
82	(C :0:3)	---	3---	(O :0:1)	5	3277891
83	(C :1:2)	---	3---	(C :1:3)	1	3230802
84	(C :0:3)	---	3---	(C :1:2)	1	3229763
85	(C :0:3)	---	3---	(C :0:3)	3	3214403
86	(C :0:2)	---	3---	(C :0:3)	1	3214402
87	(N :0:3)	---	2---	(N :1:2)	1	2312387
88	(C :1:3)	---	2---	(N :1:2)	2	2312275
89	(C :1:2)	---	2---	(N :1:2)	1	2312274
90	(C :1:2)	---	2---	(N :0:3)	1	2296914
91	(C :0:3)	---	2---	(N :0:3)	1	2296899
92	(C :1:3)	---	2---	(N :0:1)	1	2294867
93	(C :1:3)	---	2---	(O :1:1)	1	2245715
94	(C :0:3)	---	2---	(O :0:2)	2	2230339
95	(C :0:3)	---	2---	(O :0:1)	2	2229315
96	(C :0:2)	---	2---	(O :0:1)	1	2229314
97	(C :1:3)	---	2---	(C :1:3)	1	2182227
98	(C :0:3)	---	2---	(C :0:3)	3	2165827
99	(C :0:2)	---	2---	(C :0:3)	1	2165826